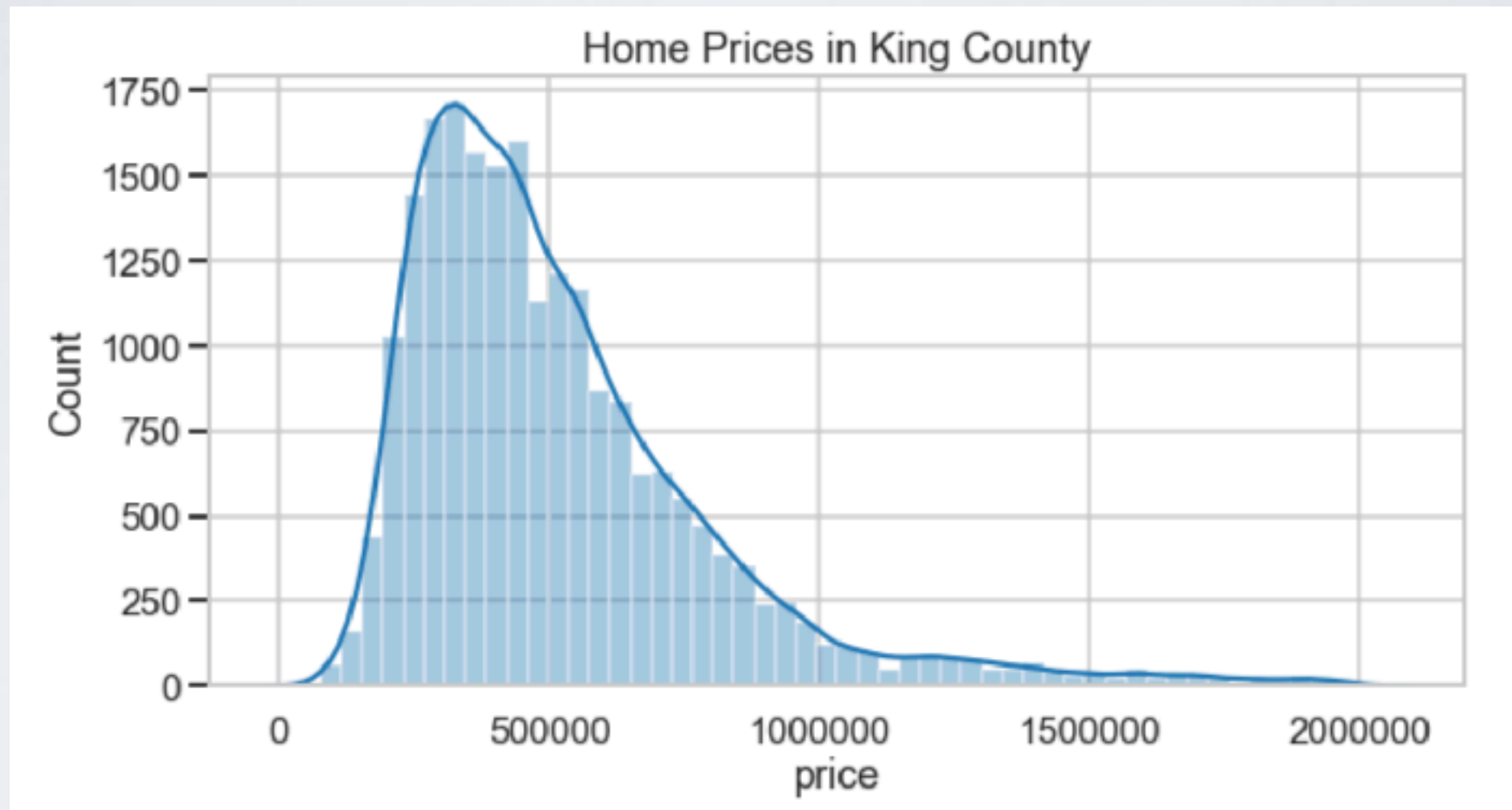# HOUSING DATA ANALYSIS

FOR HOMES IN KING COUNTY, WA

Dennis Trimarchi

2019-06-02

MOTIVATION & INITIAL ANALYSIS

- **Making predictions is useful**! Assessors make a living off of predicting home values. We want to see if an algorithm can be developed to accurately predict a home's price in King County, WA.

mean:

~ **$520K**

median:

~**$450K**

- A reliable set of data is needed: Removed 634 out of 21597 records that had missing / bad data. **Removed homes valued over $2M** because they are not comparable to the majority of homes.

GEOGRAPHY

Expensive homes in King County are *clustered* in the vicinity of: Bellevue, WA.

Having a home in this region results in increased price.

Price by Geographical Location
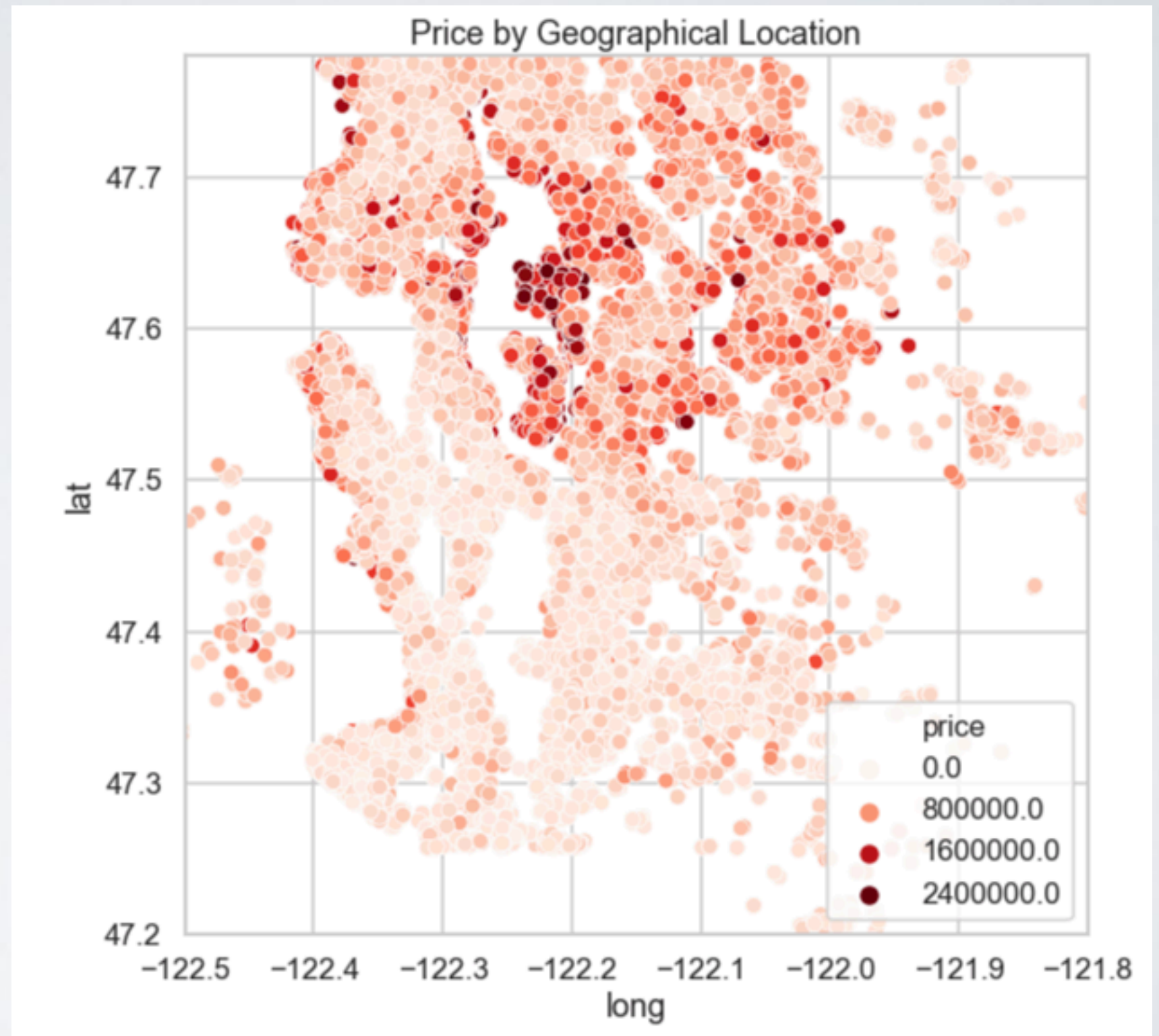
WHAT DRIVES PRICE?

Price has strongest relationship with:

Home Size (square feet living space),
Home Grade (King County grade),
**Location** (miles from Bellvue, WA)

However, Home Size and Home Grade are highly related to each other in addition to price. This is bad because predictor variables in a model need to be independent. One needs to be chosen.

HOME SIZE & HOME GRADE

Home Size and Home Grade are so similar that you can use one to predict the other. — **Higher grades are given to larger homes.**

I am getting no new information about price by including one versus the other.

LIVING SPACE VS PRICE

Price vs Home Size, color showing Location
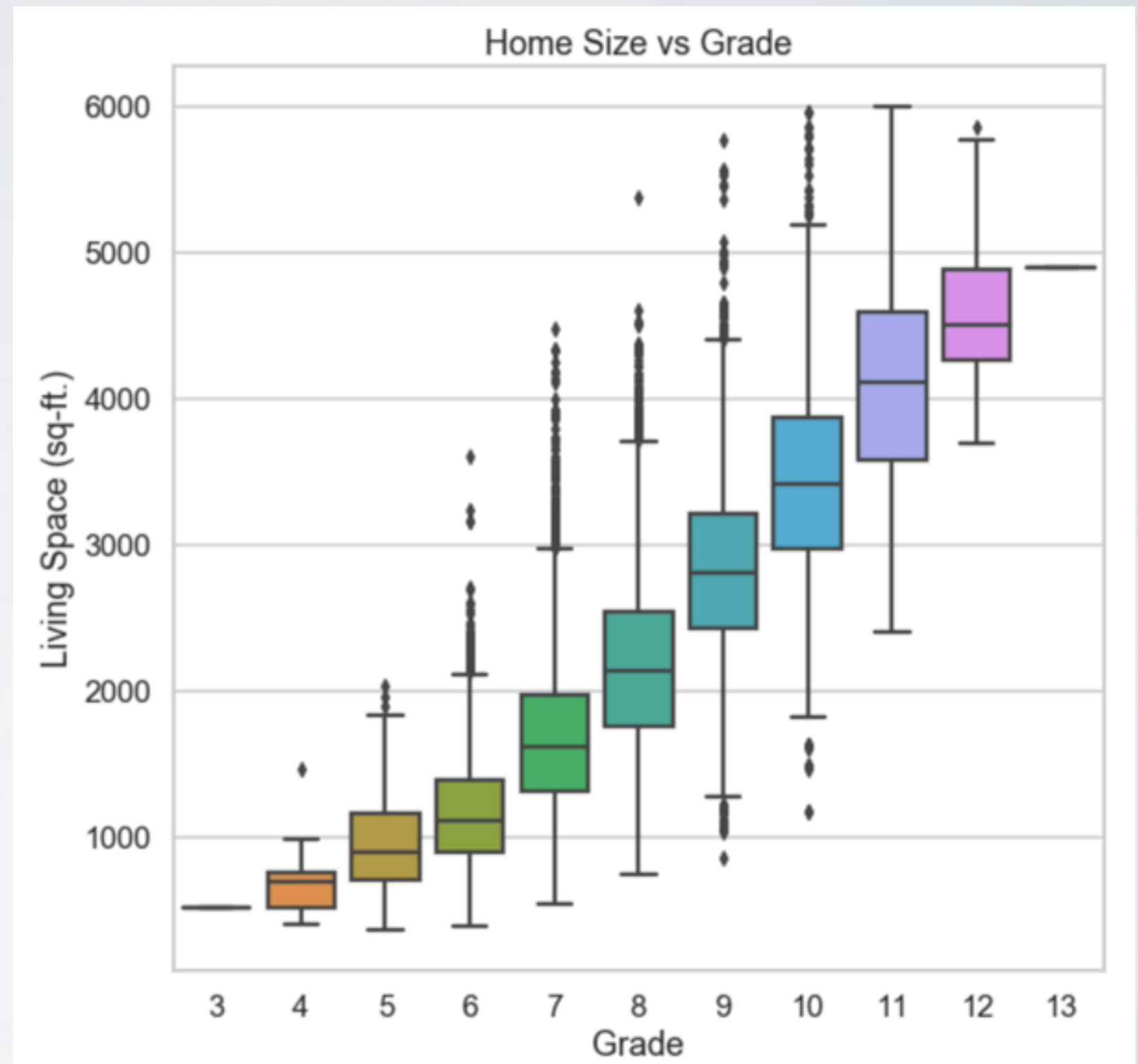
Clear positive relationship between **price and living space**.

Also, **distance from Bellevue, WA** shown through the use of color.

As you can see, homes farther away from Bellevue are clustered at the bottom of the spread, indicating lower price.

ALGORITHM & DISCUSSION

Price is based on **Home Size** and **Location**. These particular features have a strong relationship with price and independence from each other. This makes for a robust model.

$$price = 202.8 * sqft\_living - 196,800 * \ln(dist\_from\_Bellvue) + 550,300$$

The Model is simple yet effective:
- only needs two inputs
- uses readily available information: square footage, and location.

Limitations:
- It can only be applied to data for King County.
- The algorithm only explains 61% of the variation in home price for the dataset. Other factors come into play when it comes to price.

FURTHER WORK

- Some additional information gathering would improve this model.

- **Waterfront**: In the Seattle area waterfront property is everywhere, and it would be nice to see if having a waterfront view impacts price. Unfortunately, for many of the homes this information was not provided, as such it was not included in the model.

- **Home Size vs Grade**: Home Size was chosen over grade even though both features are highly related to price. With more data, it's possible that grade could be a better predictor. Also, it would be beneficial to know how the details of the King County grading system works.

- **Lot Size**: Another feature that could be investigated further is lot size. The data seemed to indicate a city vs. rural split.

THANK YOU!

- **Thank you for your time.** I hope that you've enjoyed the information presented.

- For anyone interested, there are some backup slides provided which are relevant to potential future work.

BACKUP SLIDES

WATERFRONT
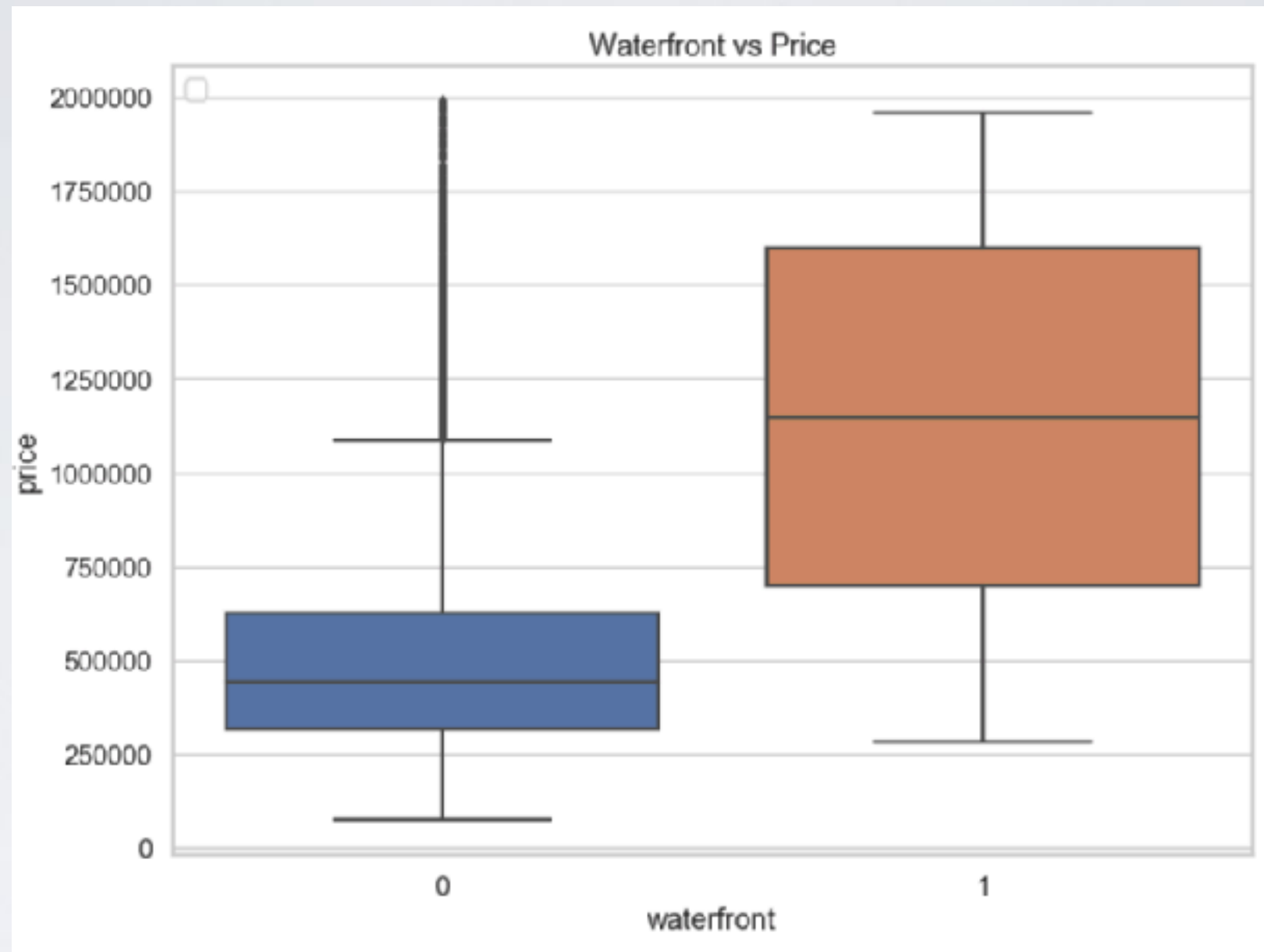
- Clear difference in price for waterfront properties. With a more complete set of data, waterfront could be added to this model.

LOT SIZE

Price vs Lot Size

- Having neighbors with large lot sizes (sqft_lot15) could indicate more rural areas. This impact on price could be investigated further.